

## CLAIMS

What is claimed is:

1. A system for extracting information comprising:  
a query input;  
a database of documents;  
a plurality of classifiers arranged in a hierarchical cascade of classifier layers,  
wherein each classifier comprises a set of weighted training data points comprising feature vectors representing any portion of a document, and wherein said classifiers are operable to retrieve documents from said database matching said query input; and  
a terminal classifier weighing an output from said cascade according to a rate of success of query terms being matched by each layer of said cascade.
2. The system of claim 1, wherein each classifier accepts an input distribution of data points and transforms said input distribution to an output distribution of said data points.
3. The system of claim 2, wherein each classifier is trained by weighing training data points at each classifier layer in said cascade by an output distribution generated by each previous classifier layer, and wherein weights of said training data points of said first classifier layer are uniform.
4. The system of claim 3, wherein each classifier is trained according to said query

input.

5. The system of claim 2, wherein said query input is based on a minimum number of example documents.

6. The system of claim 1, wherein said document comprises data points comprising feature vectors representing any portion of said document.

7. The system of claim 1, wherein said documents comprise a file format capable of being represented by said feature vectors.

8. The system of claim 1, wherein said documents comprise any of text files, images, web pages, video files, and audio files.

9. The system of claim 2, wherein a classifier at each layer in said hierarchical cascade is trained for each layer with an expectation maximization methodology that maximizes a likelihood of a joint distribution of said training data points and latent variables.

10. The system of claim 9, wherein each layer of said cascade of classifiers is trained in succession from a previous layer by said expectation maximization methodology, wherein said output distribution is used as an input distribution for a succeeding layer.

11. The system of claim 9, wherein each layer of said cascade of classifiers is trained by

successive iterations of said expectation maximization methodology until a convergence of parameter values associated with said output distribution of each layer occurs in succession.

12. The system of claim 11, wherein said successive iterations comprise a fixed number of iterations.

13. The system of claim 9, wherein all layers of said cascade of classifiers are trained by successive iterations of said expectation maximization methodology until a convergence of parameter values associated with output distributions of all layers occurs, wherein during each step of the of said iterations, the output distribution of each layer is used to weigh the input distribution of a succeeding layer.

14. The system of claim 13, wherein said successive iterations comprise a fixed number of iterations.

15. The system of claim 2, wherein each classifier layer generates a relevancy score associated with each a data point, wherein said relevancy score comprises an indication of how closely matched said data point is to said example documents.

16. The system of claim 2, wherein each classifier layer generates a relevancy score associated with said document, wherein said relevancy score is calculated from relevancy scores of individual data points within said document.

17. The system of claim 5, wherein said terminal classifier generates a relevancy score associated with each data point, wherein said relevancy score comprises an indication of how closely matched said data point is to said example documents, and wherein said relevancy score is computed by combining relevancy scores generated by classifiers at each layer of the cascade.

18. The system of claim 2, wherein said terminal classifier generates a relevancy score associated with a document, wherein said relevancy score is calculated from relevancy scores of individual data points within said document.

19. The system of claim 1, wherein features of said feature vectors comprise words within a range of words located proximate to entities of interest in said document.

20. A method of extracting information, said method comprising:

inputting a query;

searching a database of documents based on said query;

retrieving documents from said database matching said query using a plurality of classifiers arranged in a hierarchical cascade of classifier layers, wherein each classifier comprises a set of weighted training data points comprising feature vectors representing any portion of a document; and

weighing an output from said cascade according to a rate of success of query terms being matched by each layer of said cascade, wherein said weighing is performed using a terminal classifier.

21. The method of claim 20, wherein each classifier accepts an input distribution of data points and transforms said input distribution to an output distribution of said data points.
22. The method of claim 21, wherein each classifier is trained by weighing training data points at each classifier layer in said cascade by an output distribution generated by each previous classifier layer, and wherein weights of said training data points of said first classifier layer are uniform.
23. The method of claim 22, wherein each classifier is trained according to said query input.
24. The method of claim 21, wherein said query input is based on a minimum number of example documents.
25. The method of claim 20, wherein said document comprises data points comprising feature vectors representing any portion of said document.
26. The method of claim 20, wherein said documents comprise a file format capable of being represented by said feature vectors.
27. The method of claim 20, wherein said documents comprise any of text files, images, web pages, video files, and audio files.

28. The method of claim 21, wherein a classifier at each layer in said hierarchical cascade is trained for each layer with an expectation maximization methodology that maximizes a likelihood of a joint distribution of said training data points and latent variables.

29. The method of claim 28, wherein each layer of said cascade of classifiers is trained in succession from a previous layer by said expectation maximization methodology, wherein said output distribution is used as an input distribution for a succeeding layer.

30. The method of claim 28, wherein each layer of said cascade of classifiers is trained by successive iterations of said expectation maximization methodology until a convergence of parameter values associated with said output distribution of each layer occurs in succession.

31. The method of claim 30, wherein said successive iterations comprise a fixed number of iterations.

32. The method of claim 28, wherein all layers of said cascade of classifiers are trained by successive iterations of said expectation maximization methodology until a convergence of parameter values associated with output distributions of all layers occurs, wherein during each step of the of said iterations, the output distribution of each layer is used to weigh the input distribution of a succeeding layer.

33. The method of claim 32, wherein said successive iterations comprise a fixed number

of iterations.

34. The method of claim 21, wherein each classifier layer generates a relevancy score associated with each a data point, wherein said relevancy score comprises an indication of how closely matched said data point is to said example documents.

35. The method of claim 21, wherein each classifier layer generates a relevancy score associated with said document, wherein said relevancy score is calculated from relevancy scores of individual data points within said document.

36. The method of claim 24, wherein said terminal classifier generates a relevancy score associated with each data point, wherein said relevancy score comprises an indication of how closely matched said data point is to said example documents, and wherein said relevancy score is computed by combining relevancy scores generated by classifiers at each layer of the cascade.

37. The method of claim 21, wherein said terminal classifier generates a relevancy score associated with a document, wherein said relevancy score is calculated from relevancy scores of individual data points within said document.

38. The method of claim 20, wherein features of said feature vectors comprise words within a range of words located proximate to entities of interest in said document.

39. A program storage device readable by computer, tangibly embodying a program of instructions executable by said computer to perform a program storage device of extracting information, said program storage device comprising:

inputting a query;

searching a database of documents based on said query;

retrieving documents from said database matching said query using a plurality of classifiers arranged in a hierarchical cascade of classifier layers, wherein each classifier comprises a set of weighted training data points comprising feature vectors representing any portion of a document; and

weighing an output from said cascade according to a rate of success of query terms being matched by each layer of said cascade, wherein said weighing is performed using a terminal classifier.

40. The program storage device of claim 19, wherein each classifier accepts an input distribution of data points and transforms said input distribution to an output distribution of said data points.

41. The program storage device of claim 40, wherein each classifier is trained by weighing training data points at each classifier layer in said cascade by an output distribution generated by each previous classifier layer, and wherein weights of said training data points of said first classifier layer are uniform.

42. The program storage device of claim 41, wherein each classifier is trained according



to said query input.

43. The program storage device of claim 40, wherein said query input is based on a minimum number of example documents.

44. The program storage device of claim 39, wherein said document comprises data points comprising feature vectors representing any portion of said document.

45. The program storage device of claim 39, wherein said documents comprise a file format capable of being represented by said feature vectors.

46. The program storage device of claim 39, wherein said documents comprise any of text files, images, web pages, video files, and audio files.

47. The program storage device of claim 40, wherein a classifier at each layer in said hierarchical cascade is trained for each layer with an expectation maximization methodology that maximizes a likelihood of a joint distribution of said training data points and latent variables.

48. The program storage device of claim 47, wherein each layer of said cascade of classifiers is trained in succession from a previous layer by said expectation maximization methodology, wherein said output distribution is used as an input distribution for a succeeding layer.

49. The program storage device of claim 47, wherein each layer of said cascade of classifiers is trained by successive iterations of said expectation maximization methodology until a convergence of parameter values associated with said output distribution of each layer occurs in succession.

50. The program storage device of claim 49, wherein said successive iterations comprise a fixed number of iterations.

51. The program storage device of claim 47, wherein all layers of said cascade of classifiers are trained by successive iterations of said expectation maximization methodology until a convergence of parameter values associated with output distributions of all layers occurs, wherein during each step of the of said iterations, the output distribution of each layer is used to weigh the input distribution of a succeeding layer.

52. The program storage device of claim 51, wherein said successive iterations comprise a fixed number of iterations.

53. The program storage device of claim 40, wherein each classifier layer generates a relevancy score associated with each a data point, wherein said relevancy score comprises an indication of how closely matched said data point is to said example documents.

54. The program storage device of claim 40, wherein each classifier layer generates a

relevancy score associated with said document, wherein said relevancy score is calculated from relevancy scores of individual data points within said document.

55. The program storage device of claim 43, wherein said terminal classifier generates a relevancy score associated with each data point, wherein said relevancy score comprises an indication of how closely matched said data point is to said example documents, and wherein said relevancy score is computed by combining relevancy scores generated by classifiers at each layer of the cascade.

56. The program storage device of claim 40, wherein said terminal classifier generates a relevancy score associated with a document, wherein said relevancy score is calculated from relevancy scores of individual data points within said document.

57. The program storage device of claim 39, wherein features of said feature vectors comprise words within a range of words located proximate to entities of interest in said document.